# A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models

Year: 2024 | Citations: 322 | Authors: S. Tonmoy, S. M. M. Zaman, Vinija Jain, Anku Rani, Vipula Rawte

---

## Abstract

As Large Language Models (LLMs) continue to advance in their ability to write human-like text, a key challenge remains around their tendency to hallucinate generating content that appears factual but is ungrounded. This issue of hallucination is arguably the biggest hindrance to safely deploying these powerful LLMs into real-world production systems that impact people's lives. The journey toward widespread adoption of LLMs in practical settings heavily relies on addressing and mitigating hallucinations. Unlike traditional AI systems focused on limited tasks, LLMs have been exposed to vast amounts of online text data during training. While this allows them to display impressive language fluency, it also means they are capable of extrapolating information from the biases in training data, misinterpreting ambiguous prompts, or modifying the information to align superficially with the input. This becomes hugely alarming when we rely on language generation capabilities for sensitive applications, such as summarizing medical records, financial analysis reports, etc. This paper presents a comprehensive survey of over 32 techniques developed to mitigate hallucination in LLMs. Notable among these are Retrieval Augmented Generation (Lewis et al, 2021), Knowledge Retrieval (Varshney et al,2023), CoNLI (Lei et al, 2023), and CoVe (Dhuliawala et al, 2023). Furthermore, we introduce a detailed taxonomy categorizing these methods based on various parameters, such as dataset utilization, common tasks, feedback mechanisms, and retriever types. This classification helps distinguish the diverse approaches specifically designed to tackle hallucination issues in LLMs. Additionally, we analyze the challenges and limitations inherent in these techniques, providing a solid foundation for future research in addressing hallucinations and related phenomena within the realm of LLMs.